

ギターエフェクト再現に向けた音色比較手法

大豆生田 利章 *

(2019 年 11 月 27 日受理)

1 はじめに

近年，世界の音楽シーンにおいてエレキギターは欠かせない楽器となっている．その理由の 1 つに，出せる音色が多いことがあげられる．エレキギターは，弦を弾いた時の振動を電気信号に変え，アンプで音を増幅することでスピーカーから音を出している．そのため，電気信号の段階でフィルタ等の回路を通過させることにより音色を変化させることができる．このように音色を変化させることを目的とした「エフェクター」は広く普及しており，多数のエフェクターを使い分けているプロミュージシャンも多い．

アマチュアミュージシャンがプロミュージシャンの楽曲をコピーしようとした場合，仮に同じギターやアンプを揃えたとしても，同じエフェクターがなければ全く同じ音色にはならない．また，エフェクターの種類やパラメータを公開しているミュージシャンは少ないうえ，すべての機材を入手するのは経済的に難しい場合もある．そのため，音色の再現は各個人が経験と感覚をもとに試行錯誤して行っており，初心者にとっては決して簡単な作業とはいえない．

本報告では，エフェクターによって変化した音色を再現しやすくすることを目的として，複数の音色比較手法を検討する．

2 音色再現手法

有山らは図 1 に示すようなエフェクターによる音色の再現をコンピュータによって支援するシステムを提案している [1]．この手法はコンピュータ上で動作するプラグインを用い，多様化するエフェクターの接続順，パラメータの組み合わせの探索を遺伝的アルゴリズムによって効果的に行うものである．

本報告では，有山の報告をもとに図 2 のようなシステムを想定し，破線部内を自動化するための基礎となる音色比較手

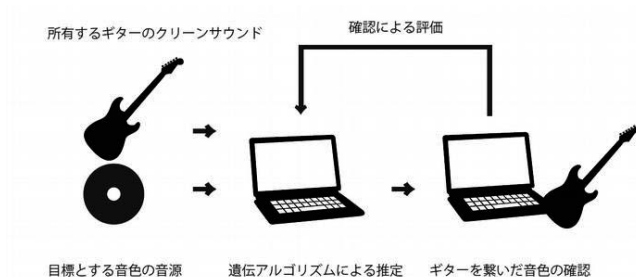


図 1 音色再現手法 ([1] による)

法の検討を行う．この部分ではシステムに必要不可欠な適合度の判定，すなわちどれだけ理想の音色に近いかどうかを評価する．具体的には，エフェクトのかかり方が異なる複数のサンプルどうしを比較し，できるだけ人間の聴覚と同じように音色の「近さ」を判定する手法を検討する．この手法により，異なるエフェクターやギターを使っている場合でも，求める音色に近い音色を再現することが可能になる．

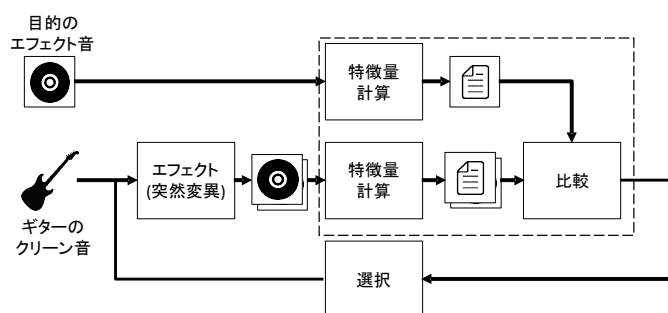


図 2 本報告の想定システム

3 特徴量

音色がどれだけ近いかを評価する基準として，後述する単純な方法で比較できる特徴量を検討対象の候補とし，それらの特性を評価する．検討対象の中から効果的な特徴量を定める条件として，次の 4 つを定める．

* 電子情報工学科

1. 人間の聴覚的特徴量との近似性
2. 実際のパラメータとの対応
3. 特徴量の計算コスト
4. 特徴量同士の比較コスト

3.1 波形データ (Direct)

波形データそのものを比較し、差の絶対値の総和を用いることによって評価する。音の長さによって評価値が変化するため、サンプル数による平均をとったものも検討対象とした。

3.2 周波数スペクトラム (FFT)

波形データをフーリエ変換してえられた周波数スペクトラムに対して、3.1 節の波形データと同様に、それぞれの周波数成分の差の絶対値を合計して評価する。長時間の音データに対しては短時間フーリエ変換 (SFFT) を用いることになるため、短時間フーリエ変換後、合計をとった全体の周波数スペクトラムの比較、および時間ごとの周波数スペクトラムすべての比較、それぞれを検討した。

3.3 メル周波数ケプストラム係数 (MFCC)

音声認識の分野で広く用いられる音の特徴量として、メル周波数ケプストラム係数（以降 MFCC とする）がある。ケプストラムとは振幅スペクトルの対数に対して再度フーリエ変換を実行したものである。対数ケプストラムに対して以下の式で表される人間の聴覚上の音の高さの指標であるメル尺度で重みづけをして MFCC を得る。

$$\text{mel} = \left\lceil \frac{1000}{\ln \left(\frac{1000}{f_0} - 1 \right)} \right\rceil \cdot \ln \left(\frac{f}{f_0} + 1 \right) \quad (1)$$

これにより、対数ケプストラムの低次成分に人の周波数知覚特性であるメル尺度で重みづけした特徴量が得られる。

MFCC は音響信号に対しても広く用いられ、その有効性は Logan らによって検証が行われている [2]。また、馬場らによる楽器音同定ではギターの音を 85% 以上の確率で認識することに成功している [3]。MFCC は音の大まかな特徴を損なうことなく、特徴量ベクトルの次元を圧縮できるという利点がある。したがって、MFCC の特徴量ベクトルの差によって音色を評価できると考えられる。

なお、本報告では時間ごとに MFCC を求めたものを SMFCC と呼ぶことにする。

3.4 メタ MFCC (meta-MFCC)

3.3 節の短時間 MFCC の各次元の時間変化を音の波形とみなし、再度 MFCC を求める。MFCC の次元数を n とすると $n \times n$ の正方行列が得られる。本報告ではこの特徴量をメ

タ MFCC (meta-MFCC) と呼ぶことにする。メタ MFCC を用いることで、より少ないデータで音色の時間的な変化もとらえられる特徴量を得ることができると考えられる。メタ MFCC の実行例を図 3 に示す。縦軸は周波数、横軸は時間である。

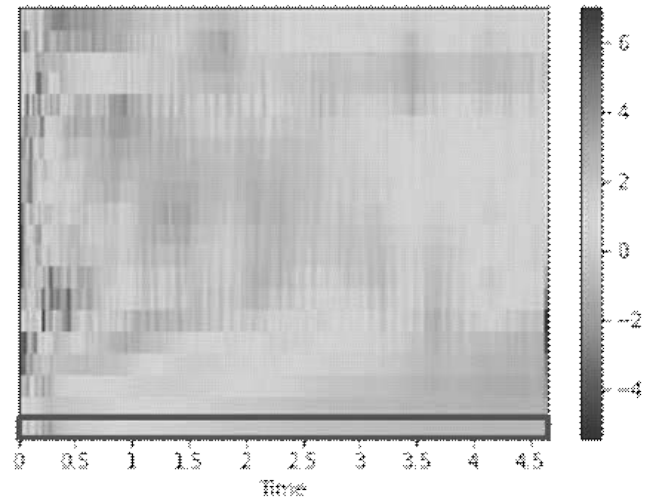


図 3 メタ MFCC の実行例

以上の処理のうち FFT と SFFT を合わせて FFT 系と呼び、MFCC, SMFCC およびメタ MFCC を総称して MFCC 系と呼ぶことにする。

4 使用ライブラリ

上記の特徴量を調査するためのライブラリとして Aquila 3.0 を用いた [4]。このライブラリはデジタル信号処理 (DSP) の機能を主にまとめている。このライブラリを用いることで、音源ファイル (.wav) から信号を得て、信号データからスペクトルを求めることができる。

5 人間によるエフェクト評価

ここでは、人間によるエフェクトを評価した結果について報告する。元となる 1 つの音源に以下に示す 3 種類 (Distortion, Reverb, Delay) のエフェクト [5] をかけたものを用意した。

Distortion 音響信号をクリップして波形をひずませ倍音を増やすことにより、温かみがある音や、ガラガラした音や、ふわふわした音を作る。

Reverb 多数の反響音を用いることで、大聖堂や音楽ホールにおける空間音をシミュレートする。

Delay 複製した音響信号を少し時間遅延させて元の音響信

号に加えることでエコーを発生させる。

各エフェクトで調節したパラメータは、以下のものである。

- Distortion : Gain
- Reverb : Size
- Delay : Feedback

まず、それぞれのエフェクトの強弱を 0% から 100% の間で 25% ずつ強くし、5 段階の音源を用意した。これらのエフェクトを与えた音源に対して順序尺度による評価を行う。ここでは、3 種類のエフェクトの強弱をそれぞれ聞き分けて、強く感じた順に 5 つの音源を並び替えるアンケートを 10 人に実施した。表 1 に結果を示す。

表 1 人間による評価の平均

実順位	人間の評価した順位		
	Distortion	Reverb	Delay
1	1.33	1.67	2.25
2	1.92	2.25	2.25
3	3.17	3.58	4.25
4	4.17	3.42	2.42
5	4.42	4.08	3.83

表 1 の結果は、各音源につけられた順位を平均したものである。Distortion はほぼ実順位に近い値を得た。このことから、エフェクトの種類によっては強弱をほぼパラメータの違い通りに聞き分けることができるといえる。一方、Reverb はある程度しか強弱を判断できておらず、Delay はほとんど判断できていないことが分かった。図 4 に Distortion に関して、実順位と人間の評価した順位の関係をグラフの形で示す。

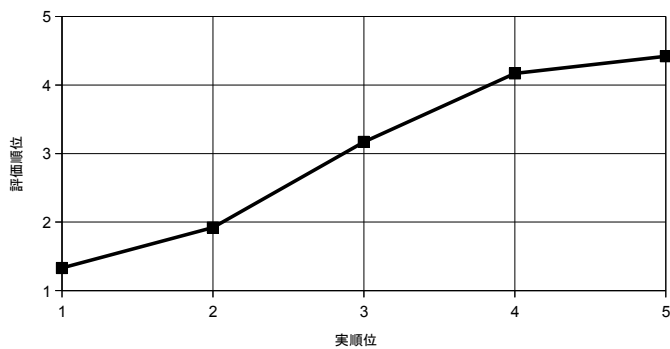


図 4 Distortion に対する実順位と人間の評価順位

6 特徴量による比較実験

6.1 実験条件

第 3 節で挙げた direct, FFT, MFCC を用いて、音源から特徴量を抽出する実験を行った。第 5 節で用いたエフェクトのパラメータを 10% 刻みで 0% から 100% まで変えた音源を使用した。本実験の特徴量計算の条件を表 2 に示す。

表 2 特徴量の計算条件

フレーム長	1024 サンプル
オーバーラップ幅	50%
MFCC 次元数	12 次元
MFCC フィルタバンク数	24
MFCC リフタ長	22

6.2 実験結果

第 5 節の人間対象の評価実験と同様の音源を用いて、音色を比較するための特徴量の抽出と評価を行った。特徴量同士の差の合計値よりマンハッタン距離を求め、距離の近い順に順位付けを行うことで評価をした。各比較結果は、エフェクトをかけていないものを基準として、最も距離の差が大きいものが 1.0 となるよう正規化した。比較結果を表 3 から表 5 に示す。

表 3 は Distortion の強弱を比較した結果を示している。SFFT が実際のパラメータと誤差 $\pm 3\%$ 以内で近いことがわかった。ただし、図 4 に示したように、人間による評価ではパラメータが 50% を超えた領域では評価値が飽和していく。この特性に最も近い評価になったのはメタ MFCC である。図 5 にメタ MFCC を用いた特徴量抽出の結果のうち、Distortion に関するものを示す。図 4 に示した人間による評価に近い評価ができていることが分かる。

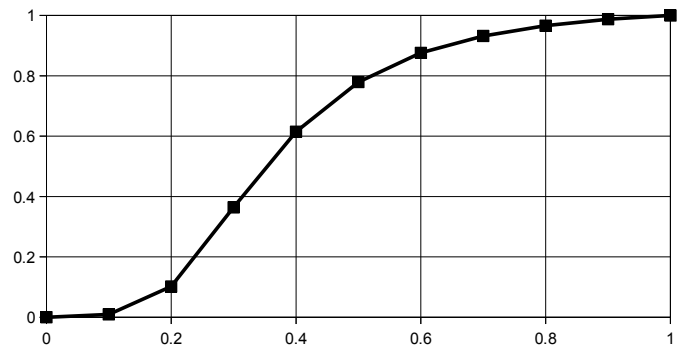


図 5 メタ MFCC を用いた特徴量抽出 (distortion)

表 3 Distortion における各特微量での比較結果

parameter	difference					
	direct	FFT	SFFT	MFCC	SMFCC	meta-MFCC
0%	0	0	0	0	0	0
25%	0.2549	0.3687	0.2217	0.2969	0.2808	0.4663
50%	0.5664	0.4420	0.5249	0.6359	0.6062	0.8491
75%	0.8064	0.7509	0.7774	0.8459	0.8293	0.9709
100%	1	1	1	1	1	1

表 4 Reverb における各特微量での比較結果

parameter	difference					
	direct	FFT	SFFT	MFCC	SMFCC	meta-MFCC
0%	0	0	0	0	0	0
25%	0.4287	0.4174	0.3890	0.2279	0.3861	0.3635
50%	0.7050	0.6672	0.6609	0.5595	0.6682	0.6676
75%	0.8733	0.8459	0.8514	0.7988	0.8513	0.8461
100%	1	1	1	1	1	1

表 5 Delay における各特微量での比較結果

parameter	difference					
	direct	FFT	SFFT	MFCC	SMFCC	meta-MFCC
0%	0	0	0	0	0	0
25%	0.2342	0.2070	0.1864	0.3494	0.2352	0.2602
50%	0.4661	0.4288	0.3889	0.6182	0.4615	0.5205
75%	0.6946	0.6750	0.6292	0.7812	0.6865	0.7235
100%	1	1	1	1	1	1

表 4 は Reverb の強弱を比較した結果を示している。MFCC を除き、どの比較手法も実際のパラメータより +10% 以上も高い数値を出した。

表 5 は Delay の強弱を比較した結果を示している。ここではメタ MFCC が $\pm 2\%$ 程度の誤差でかなり近い値となった。また、Direct も大きくて -5% 程度の誤差でパラメータを推定することができた。

6.3 実行速度の測定

各特微量を比較処理する速度を、サンプル数 (音源の長さ) を変えながら測定した。サンプル数は 1.25×10^6 , 2.5×10^6 , 5.0×10^6 , 1.0×10^7 の 4 通りである。測定結果を図 6 に示す。どの手法でも、データ数 N に対する計算に要する時間は $O(N)$ であったが、FFT、MFCC、メタ MFCC はサンプル数 1.0×10^7 (一般的な音声ファイルでは約 3 分半のデータ量) でもほぼ 0.1 ms で実行を完了することができた。一

方、SFFT による求め方ではサンプル数 1.0×10^7 でおよそ 10.0 s と、音色同士の比較にはかなりの時間を要した。

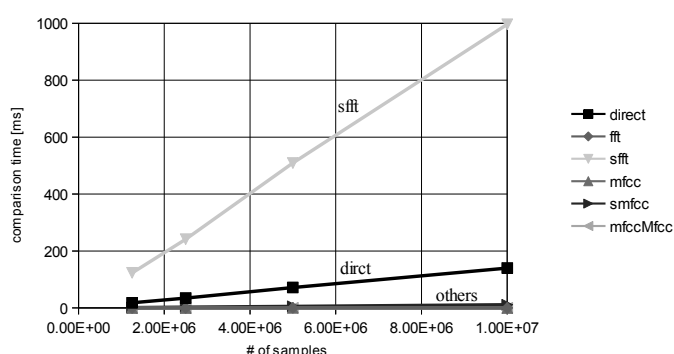


図 6 各特微量の比較処理時間

図 6 を拡大して、FFT、MFCC、SMFCC、メタ MFCC の比較処理時間を比較したものが図 7 である。SMFCC は

若干処理時間がかかっているが、FFT、MTCC およびメタ MFCC の処理時間はほとんど同じになった。

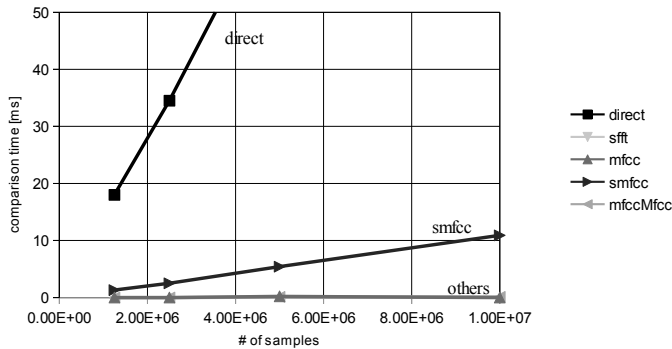


図7 各特徴量の比較処理時間（拡大図）

次に、FFT と SFFT によるスペクトラム、MFCC、SM-FCC を算出するまでの実行時間を測定した結果を図8に示す。どの特徴量に対しても計算時間は $O(N)$ であった。MFCC 系3種の計算時間はFFT系2種に比べて5倍近く増加することが判明した。また、MFCC系およびFFT系同士では計算コストの差はほぼなかった。

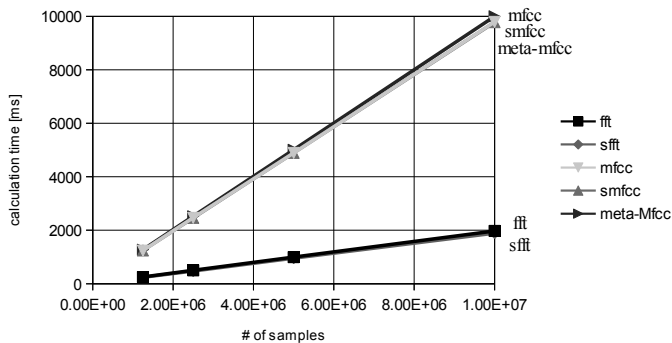


図8 各特徴量の計算時間

7 まとめ

コンピュータで特徴量を用いて計算した場合と人の聴力による評価と比べた。その結果、Distortion ではパラメータはどの特徴量でもかなり正確に推定できることが判明した。実際のパラメータとの対応では、3種のエフェクトを通じてMFCC系が良好であると判明した。その中でもSMFCCとメタMFCCが特に良いことがわかった。計算コストはMFCC系に比べるとFFT系は20%程度に抑えられた。またFFT系あるいはMFCC系内ではコストの差はほぼなかった。比較にかかる時間コストはFFTとMFCC系が低く、その中でもFFT、MFCC、およびメタMFCCが特にコストが抑えられていた。以上に述べたことから、エフェ

クトを加えた音色の判定条件である人間の聴覚的特徴にとの近似性、比較にかかる時間コスト、特徴量の計算にかかる時間コストの順に比重を置くと、メタMFCCが最適であると考えられる。

残された課題として、MFCCのパラメータや正規化処理について、その最適値を検証する必要がある。もし、最適値が現状と異なることが分かれば、今後より正確にエフェクトの強弱を判断できる手法となり得る。また、特徴量による比較手法の検証では0%～100%を5段階で検証したが、より段階数を増やして検証することも必要であると考えられる。発展課題として、複数のエフェクトのパラメータを同時に変化させた場合の2次元、3次元的な変化についての実験が考えられる。

謝辞

本報告では、平成30年度卒業研究生の室井隆成君（現ローランド）および町田拓斗君（現東京工科大学学部生）によるデータ収集および解析の結果を利用した。

参考文献

- [1] 有山大地, 安藤大地他, “進化論的計算を用いたエレキギターの音色再現手法の提案,” 情報処理学会インタラクショナル2016論文集, p.250, 2016/3/2.
- [2] Logan Beth, “Mel Frequency Cepstral Coefficients for Music Modeling,” Proc. of ISMIR, 2000. (http://ismir2000.ismir.net/papers/logan_paper.pdf (2019/11/26)).
- [3] 馬場貴之, 山田武志他, “HMMとMFCCを用いた楽器音の音源同定の検討,” 情報処理学会研究報告MUS-61, p.79, 2005/8/5.
- [4] Aquila 3.0, Open source DSP library for C++, <https://aquila-dsp.org/> (2019/11/26).
- [5] Wikipedia, “Effects unit”, https://en.wikipedia.org/wiki/Effects_unit (2019/11/26).

Timbre Comparison Method to Reproduce Guitar Effect

Toshiaki OHMAMEUDA

In recent years, electric guitars have become an indispensable instrument in the world music scene. If an electric guitar is used, the timbre can be changed by passing through a circuit such as a filter at the electrical signal stage. In this way, effectors are widely used to change the timbre, and many professional musicians use a variety of effectors. When an amateur musician tries to copy the music of a professional musician, even if the same guitar or amp is prepared, the same timbre will not be achieved without the same effector. In addition, few musicians disclose the types and parameters of effectors, and the equipments are expensive. The reproduction of the timbre is done by trial and error based on experience and feeling by each musician. This is not an easy task for beginners. In this paper the timbre comparison methods is compared to easily reproduce the sound modulated by effectors. The methods evaluated in this paper are direct, FFT (fast Fourier transform), SFFT (short-time Fourier transform), MFCC (Mel-Frequency Cepstral Coefficients) and meta-MFCC. The effects evaluated in this paper are distortion, reverb and delay.

Comparing the evaluation of characteristics calculate by a computer and that by human hearing, it is shown that distortion can estimate parameters with any characteristics accurately. The MFCC system was found to be good for three effects to reproduce the original parameters, especially SMFCC (short-time Mel-Frequency Cepstral Coefficients) and meta MFCC were found to be particularly good. The computational cost of the FFT system was reduced to about 20 percent compared to the MFCC system. There was almost no difference in cost in the FFT or MFCC system. The time cost for the comparison was low in the FFT and MFCC systems, especially the FFT, MFCC, and meta MFCC. Therefore meta MFCC is considered to be optimal if the specific gravity is compared in order of the human auditory characteristics.